UNDERSTANDING LIGHTNESS AND BRIGHTNESS PERCEPTION IN REAL AND VIRTUAL MEDIA

JAYKISHAN Y. PATEL

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

December 2021

Abstract

This thesis aimed to study how lightness and brightness perception relate to each other. We used a simple task to study whether observers perceive lightness and brightness to be different percepts and what cues to they use to make these judgements. In Experiment 1, we used a custom-built apparatus to present two reflectance patches, each with independent illuminance. In the lightness and brightness conditions, observers judged which patch had a higher reflectance or luminance, respectively. In Experiment 2, we repeated the same procedure using a computer rendering of the apparatus on a monitor. Finally, we simulated computational models of lightness and brightness to evaluate their performance with respect to observer performance. We conclude that (a) lightness and brightness judgements are more similar than expected from previous work, (b) brightness is nothing like an estimate of luminance, and (c) current computational models can fail on even simple lightness and brightness judgements.

Acknowledgements

I would like first give my heartfelt thank you to my supervisor and mentor Dr. Richard Murray for his continued support, guidance and patience throughout my master's thesis. Your unending patience with the circumstances surrounding the thesis made this work possible. Working with you for the past two years has been the most creatively fulfilling and enriching experience in my academic career. I am thrilled to get started with my doctoral work soon!

I would also like to extend my thanks for Dr. Peter Kohler for his patience and guidance with the time line for the thesis. Your insightful and thorough eye has made this thesis exponentially better! I look forward to working with you in the future! Thank you to Dr. Robert Philip Chalmers and Dr Michael Brown for being a part of my committee and for providing me with an encouraging and kind environment for my defence.

Finally, I would like to acknowledge the efforts of my family in helping, guiding and supporting me thorough this degree. No amount of appreciation would be enough to fully capture the impact you all have had on my life and I am truly gifted to be surrounded by such brilliant, kind, insightful and supportive people!

A CIHR Frederick Banting and Charles Best, Ontario graduate scholarship, Vision: science to applications and a CVR Ian Howard memorial scholarship made this work possible. Thank you to the funding agencies for funding my research.

Table of Contents

List of Tables

List of Figures

1 Introduction

People have the ability to perceive the colour of objects correctly despite substantial changes in illumination (Gilchrist, 2006). This ability to estimate black, white and shades of grey under different lighting conditions by discounting lighting factors is known as lightness constancy. For instance, lightness constancy allows us to perceive a white piece of paper as white regardless of bright or dim lighting conditions. This allows us to correctly identify objects under wildly different lighting conditions. How the brain accomplishes lightness constancy is one of the biggest questions in vision science.

1.1 Key Terminology and Concepts

To be able to understand the complexity of the topic it is important to understand some key terms regarding electromagnetic radiation. Radiant flux describes the quantity of energy that is interacting with a surface over unit time and is measured in watts. Radiant intensity describes the amount of radiant flux present over a solid angle and is measured in watts/steradians. Irradiance is the amount of radiant flux present over a given area and is measured in watt/m^2. Radiance describes the amount of radiant flux present over a solid angle and area and is measured by watt/steradians $\cdot$ m^2. These are the basic terms that define how electromagnetic radiation interacts with the environment. These *radiometic* terms are used when discussing radiation at all wavelengths.

For the purposes of this thesis, we are interested in electromagnetic radiation within the human visible spectrum. This range includes includes radiation with the wavelength of about 380 to 770 nm and is referred to as visible light. All the previously mentioned terms have corresponding *photometric* terms that are only used for visible light. The photometric analogs of radiant flux, radiant intensity, irradiance and radiance are luminous flux, luminous intensity, illuminance and luminance respectively. These measurements do not use watts, instead using *lumens* which only measures the power output of visible light. The transformation radiometric measures to photometric measures is defined by the equation,

$$P = 683 \int_{380}^{770} R \cdot V(\lambda) d\lambda \tag{1}$$

Where P is any photometric measure, R is the corresponding radiometric measure, V(λ) is the photopic spectral luminous efficiency curve shown in figure 1, and λ is the wavelength of light.

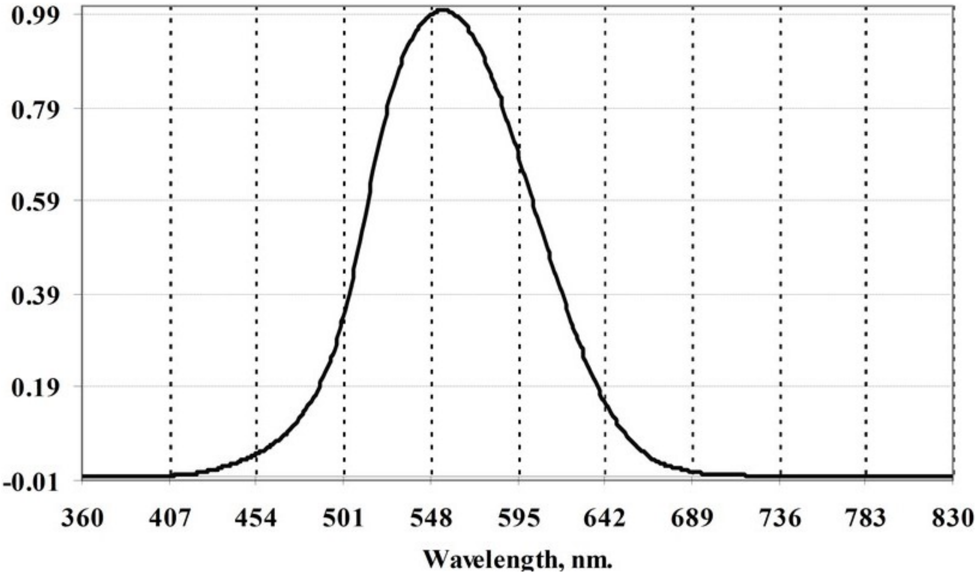


Figure 1. Photopic spectral luminous efficiency function for a CIE 1924 standard photometric observer

Visible light interactions in any environment can be defined using three components: *illumination, luminance*, and *reflectance*. Figure 2 shows how these components interact with each other. The next section will go into further detail about how this takes place.

Illumination as mentioned before refers to the amount of visible light present over over a given area. Since this is the photometric analog of irradiance, it is measured in lumens/m^2.

Luminance refers to the amount of light that is reflected off an object. Luminance is the analog of radiance and should therefore measured in lumens/m^2·steradians. However, since lumens/steradians is the same as a candela (cd), Luminance is measured in cd/m^2.

Reflectance refers to the proportion of incident light an object reflects. This means that reflectance values are measured between 0 and 1, 0 being no light reflected and 1 being all light reflected. Although the range of reflectance spans from 0 to 1, most of the observed reflectance values in natural scenes range between 0.03 to 0.90.

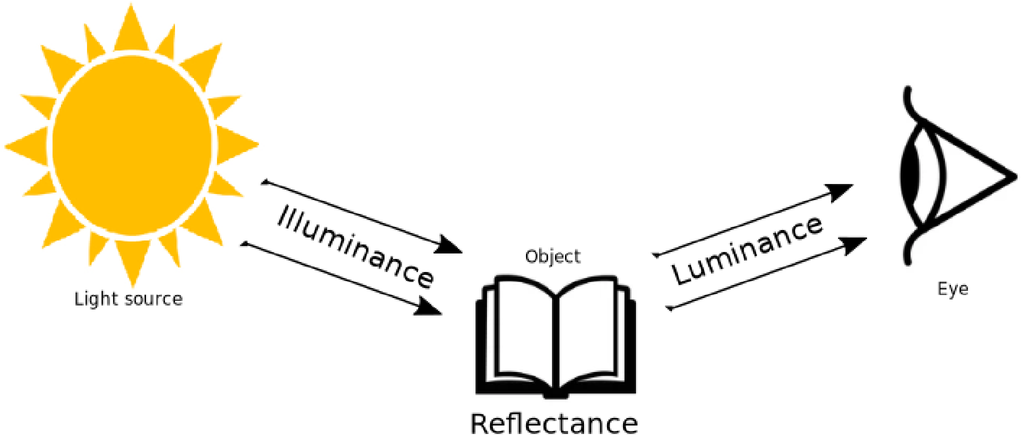


Figure 2. Light interaction explained using illumination, reflectance and luminance. A light source, in this case the sun, is the illumination. An object, in this case the book, has a reflectance factor that determines how much light is reflected and how much is absorbed. The remaining light value after absorption is the luminance captured by the eye.

Until now we have discussed the physics of how light interacts with objects and the environment. However, to understand how people perform lightness and brightness tasks it is important to discuss how light is perceived in the brain. Each of the three components of light discussed above have a perceptual analog. *Brightness* is the psychological phenomenon defined as the perceived luminance of an object. *Lightness*, on the other hand, is the perceived reflectance of an object (Adelson, 1993). Finally, *perceived illumination* is the psychological phenomenon defined as the perceived illumination present in a given scene. These measures are subjective to the observer. In other words, different observers can have different brightness estimates of an object with the same luminance and different lightness estimates for an object with the same reflectance (Brainard & Hurlbert, 2015). The retina only has access to the luminance information as the incident light is a combination of the reflectance of objects and the illumination present in the environment. As such, the brain needs to be able to extract reflectance information from this luminance information to be able to perform lightness judgements. It is not well understood how the brain accomplishes this extraction. This extraction involves separating luminance information into reflectance and illumination information without any information about the ratio between them. This means it is ambiguous which combination of reflectance and illumination creates a particular luminance. The next section delves into the ambiguity of separating luminance information into accurate predictions of reflectance and illumination.

1.2 The Ambiguity Problem

Lightness constancy requires one to be able to parse luminance information into reflectance information and illumination information. This task is more complicated than it seems however, as luminance values are proportional to products of reflectance and illumination values. To understand this ambiguity it is important to explain how luminance values are calculated. Equation 2 explains how reflectance and illumination values are used to calculate luminance. Here, L refers to the luminance value, R refers to the reflectance of the object and I refers to the illumination in the scene. An important aspect of the equation is that the same luminance value can be reached with many different combinations of reflectance and illumination. This leads to the problem of ambiguity, as there are a large number of possible scene configurations that will have the same luminance value but vastly different reflectance values.

$$L = \frac{R \times I}{\pi} \tag{2}$$

For example, a luminance value of $50/\pi\ cd/m^2$ can be achieved with a reflectance of 0.5 and illumination of 100 lx, as well as with a reflectance of 0.25 and illumination 200 lx. Hence, without knowing the ratio between the illumination and reflectance for an object in a scene performing lightness constancy accurately would be extremely difficult.

1.3 Lightness and Brightness theories

This goal of this thesis is to improve the current understanding of how people compute lightness and brightness. Therefore, a logical first step is to have an overarching understanding of how the current quantitative theories model these phenomena. To this end, a review of tested models and theories of lightness and brightness provides the information necessary to understand the current quantitative methods for measuring lightness and brightness. This section will review the anchoring theory (Gilchrist et al., 1999), Oriented Difference of Gaussian (ODOG) (Blakeslee & McCourt, 1999), the high-pass model (Shapiro & Lu, 2011), the retinex models (McCann, 1999; Frankle & McCann, 1983), and an adaptive filtering model by (Dakin & Bex, 2003).

1.3.1 Anchoring Theory.

One of the models used to explain lightness perception is known as *anchoring theory* (Gilchrist et al., 1999). One argument the anchoring theory makes is that the perceived luminance of objects is not only dependent on the object luminance but also the intensity of the surrounding environmental luminance. For example, a test patch in an experiment will appear to be lighter when it is surrounded by a dark background and vice versa (Heinemann, 1955). The model takes it a step further however, as it suggests that the visual system separates images into *frameworks* using features such as shadow boundaries and depth discontinuities. These frameworks are subsections of the image with uniform lighting. The visual system separates the image into two different types of frameworks, global and local. The global framework represents the entire image as a whole, while the local frameworks represent subsections of the image. The next step is to find the highest luminance in each local framework and using it to compute the reflectance within the framework. The equation for the computation is shown in equation 3, where R is the calculated reflectance, L is the luminance of the local framework and L_{max} is the local maximum luminance. These are the local reflectance values from the local frameworks. Next, global framework luminance is used in place of L_{max} to obtain different reflectance values for each framework. These new reflectance values are the global reflectance values as they are based on the global maximum luminance instead of the local maximum luminance. This yields two estimates of reflectance for each pixel in the image. The first is the reflectance estimate calculated using the luminance of the pixel and the maximum luminance of the *local* framework. The second is the reflectance estimate calculated using the luminance of the pixel and the maximum luminance of the *global* framework. The reflectance values computed for each image patch in the local and global frameworks are likely to be different from each other and therefore the visual system takes the weighted average of these values to compute perceived reflectance. The weight for each framework depends on how strongly the local framework is perceptually segregated from the global framework. The larger the segmentation the higher the weight of the local estimate.

$$R = 0.90 \cdot L/L_{max} \tag{3}$$

Figure 3 shows how anchoring can provide an explanation for some lightness illusions in human perception. The simultaneous contrast illusion shows that local contrast has an effect on lightness estimates. The anchoring theory explains this using the global and local framework paradigm. Luminance ratios are measured for both patches and their respective backgrounds (local frameworks) and between the patches and the global highest luminance (global framework). Since the average of these ratios is used to determine the lightness estimate the estimate of the patch on the lighter side remains unchanged as it has the same local and global contrast, while the estimate for the patch on the darker side becomes higher due to the local ratio being higher than the global ratio. This accounts for the patch appearing lighter even though it has the same physical reflectance. In terms of equation 3, the local reflectance estimate of the patch on the darker side will be close to 0.90 since the patch is has the highest luminance in the framework. The local reflectance estimate for the patch on the lighter side will be lower since the background has a much higher luminance than the patch. Finally, the global reflectance estimate of both patches will be the same but crucially, since the reflectance estimate is the weighted average of the local and global estimates, the patch on the darker side will have the higher reflectance estimate. This observed result is consistent even when the backgrounds are flipped and thus so are the estimates.

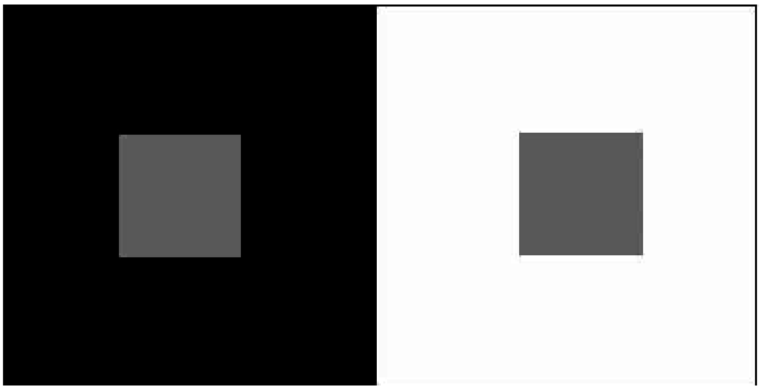


Figure 3. The simultaneous contrast illusion. Both grey patches have the same physical reflectance, but the patch on the darker background appears to be brighter than the patch on lighter background. In other words, observer lightness estimates are affected by the background on which the patch appears.

1.3.2 Oriented Difference of Gaussian (ODOG).

One of the models used to explain brightness perception is the ODOG model (Blakeslee & McCourt, 1999). The model makes use of on-center off-surround filters placed at different orientations to represent an image. These are modified versions of the on-center off-surround receptive fields generated by the ganglion cells present in the eyes. The on-center filter is circular while the off-surround filter is elliptical. The filters are oriented differently as well as tuned for different spatial frequencies. The model works by placing the filters on any image and returning a response for each pixel based on the on-center, off-surround mechanism. Figure 4 is taken from Blakeslee & McCourt (1999) and shows how the filters are used to calculate a brightness estimate. The combination of different orientations and tunings allows the filters to detect changes in luminance in a similar way to receptive fields in the visual system. The filters respond to luminance levels, hence, a high luminance in the on-center area will generate a positive response while a high luminance in the off-surround area will generate a negative response and vice versa. These filters are evenly spaced at different orientations to capture the image properties that are relevant to the brightness estimation (Figure 4b and 4d). In this way, the model attempts to capture spatial frequency differences within the image as well as the regional luminance differences in an image to provide an accurate brightness estimate. The receptive fields placed on the image at different orientations all generate a response to the image (Figure 4g). The model then normalizes the responses from these filters, in other words it bumps up the output from the filters with the lowest response to the image and lowers the response from the filters that had the highest response to the image. In this way the model generates a representation of the image by using the sum of all the filtered outputs (Figure 4h). Using this normalization and bumping of the low responses the model is able to predict brightness illusions such as the one shown.

1.3.3 Retinex. The retinex model (Land & McCann, 1971) is a model of lightness perception. This model builds on the idea of luminance ratios. The idea states that luminance perception is based on the contrast of a particular patch and its background (Heinemann, 1955). The retinex model uses this principle to generate a relative measure of the luminance present in all parts of the image. The model accomplishes this by taking the ratios of neighbouring luminances and comparing them. The model then changes all near unity luminance ratios and reduces them to 1.0. This has the effect of factoring out luminance changes that are present due to slow illumination changes while keeping intact luminance changes that occur due to reflectance. Each section of the image is then divided into regions based on its luminance. Then the model calculates the products of luminance ratios of neighbouring regions across the image and locates the region with the highest luminance. The region with the highest ratio is the highest luminance region in the

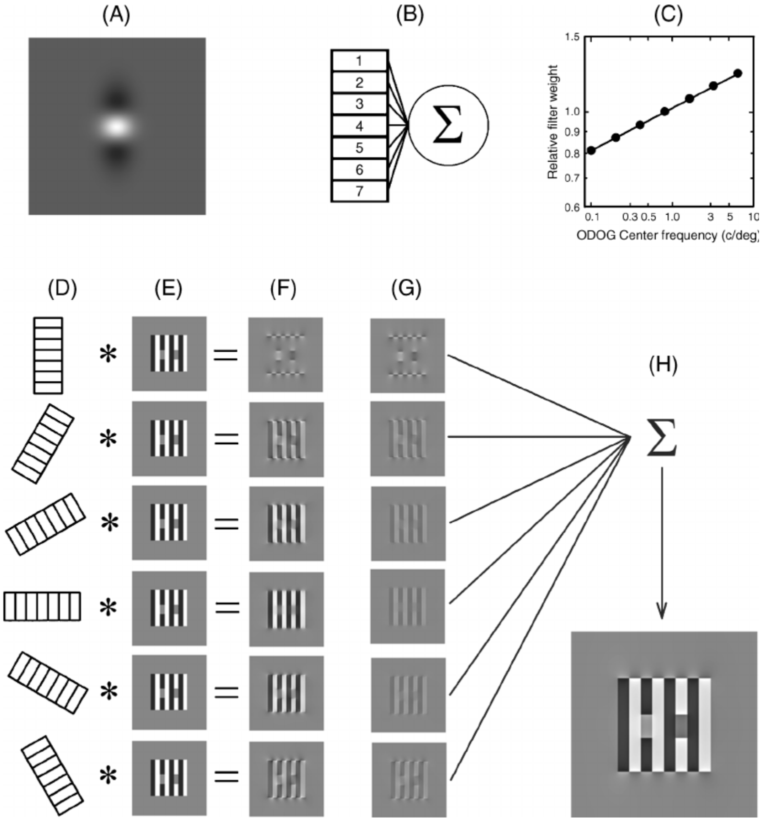


Figure 4. The ODOG brightness model. A shows the on-center, off-surround filters used by the model. B shows how the filters are arranged and then summed together over a section of the image. Panel C shows the weighting of each filter with respect to the center frequency of the filters. D - F show how the different orientations of the filters when applied to the original image return different response images. These images are then normalized as shown in G and summed to generate the model output as shown in H. The model response is able to replicate the illusion with the grey patch on the dark strip appearing lighter than the one on the white strip (Blakeslee & McCourt, 1999).

image. This highest luminance is then used as a base to normalize all other luminances present in the image. Since these luminance changes discount changes due to illumination, the remaining changes in luminance are due to changes in reflectance. This allows for the calculation of the reflectance of patches by taking the ratio of the patch luminance and dividing it by the highest luminance present in the image.

1.3.4 High pass Filter. The high pass filter model (Shapiro & Lu, 2011) is a spatial filtering model for brightness perception. The model is similar to ODOG as it also makes use of centre-surround receptive fields. Using the on-center off-surround receptive fields, the visual system is able to pinpoint the sharp changes in an image. The center of the receptive field will produce a positive response and the surrounding areas will be inhibited creating a negative response. This means that at the area of luminance change, there will be a peak in response followed by a trough. The model filter works in a very similar way to how the visual system uses lateral inhibition. The models uses on-center off-surround receptive fields and places them on each pixel of the target image. This allows for the model to get information about the luminance boundaries present in the image. This is possible since at luminance edges receptive field response will have a similar peak and trough due to the center-surround receptive fields. Furthermore, the response of the receptive fields with also reverse as the luminance changes from low to high or vice versa.

1.3.5 Adaptive Filtering Model. The adaptive filtering model (Dakin & Bex, 2003) is spatial filtering model. Natural images have a property where the amplitude of spatial frequencies follow a structure of $amplitude(f) = c * (1/f^{\alpha})$, where f is the spatial frequency, c is a constant, and α represents a negative slope in log on log coordinates and varies between 0.7-1.5 in natural images (Dakin & Bex, 2003; Field, 1987). This means that low spatial frequencies typically have higher amplitudes in natural images than high spatial frequencies. When the model is presented with an image, it applies the 1/f amplitude rule to the spatial frequency spectrum, changing the image frequencies' amplitude to match those typically present in natural images. This modified version of the image is able to accurately simulate many lightness/brightness illusions in human perception. Hence, the idea is that the visual system has a built-in mechanism to make all image frequencies have amplitudes proportional to the 1/f. An example of this in action is presented in Figure 5.

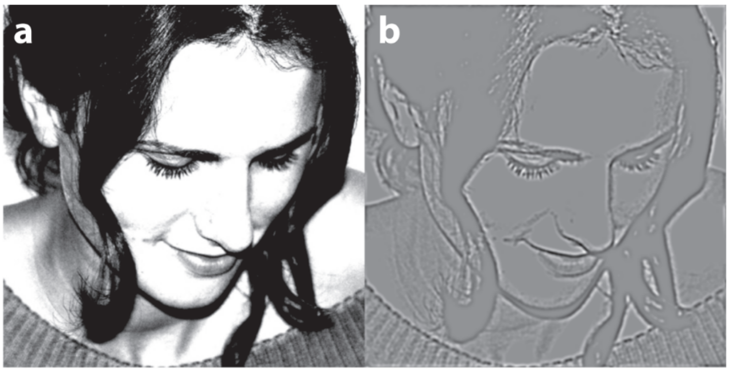


Figure 5. Excerpt image from Dakin & Bex (2003). Panel (a) is an example of a natural image. Panel (b) is the output for the first image when put through a high-pass filter. Their filtering model is able to correctly predict that the hair in the filtered image will appear darker than the face even though they have the same physical luminance.

1.4 Overview

There have been many experiments studying how lightness judgements are affected when reflectance and illumination are changed (Patel et al., 2018; Murray, 2021). However, very little is understood about how brightness judgements change with systematic changes to lighting, reflectance, and luminance. Experiment 1 in the thesis addresses this gap by measuring a *decision space* of observer responses from a physical lightness experiment and showing how lightness and brightness estimates change with reflectance and illumination. A decision space is a useful experimental tool that shows how people make judgements over a space covered by two or more controlled variables (Figure 6) Pritchett & Murray (2015). Decision spaces are further explained in a later part of the thesis. The experiment will measure how lightness and brightness judgements change by measuring a decision space for observer lightness and brightness estimates. In this instance, it will show observers' lightness and brightness judgements over a range of values of reflectance, lighting, and luminance. Comparing observer decision spaces will show how lightness and brightness relate to reflectance and illumination, and whether they indeed measure perceived reflectance and perceived luminance, respectively. Understanding how people make lightness and brightness judgements will help to decode how the brain accomplishes lightness constancy.

Flat-panel displays have been a popular medium for experiments on lightness and brightness since the 1980s. These allow for easier creation of novel stimuli and more precise control over the presented color than physically built stimuli. However, there is evidence that people's lightness and brightness judgements are not the same when using these different media (Snyder et al., 2005; Menshikova et al., 2013; Ripamonti et al., 2004; Bloj et al., 1999; Boyaci et al., 2003, 2004, 2006; Morgenstern et al., 2014; Patel et al., 2018). Experiment 2 in the thesis focuses on studying lightness and brightness judgements on a flat-panel display with recreated stimuli from the physical scene. We measured decision spaces using stimuli on the flat-panel displays. Using these decision spaces and comparing them to those from Experiment 1 allowed for an analysis of the the effect of presentation media on lightness and brightness judgements. We studied the observer responses for these experiments to gather whether lightness or brightness judgements change based on the how the stimuli are presented.

Decision spaces also provide a novel method of testing current computational models of lightness and brightness. To test model validity, the results from the decision spaces in Experiments 1 and 2 were compared to the decision spaces measured using the models. The models used for comparison include the oriented difference of Gaussians (ODOG) model (Blakeslee & McCourt, 1999), a high-pass model (Shapiro & Lu, 2011), two versions of the retinex model (McCann, 1999; Frankle & McCann, 1983), and an adaptive filtering model (Dakin & Bex, 2003). These models provided another important test group, alongside the human observers, to improve the current understanding of the computations behind lightness and brightness judgements. To this end, we test the response of the ODOG, retinex, high-pass filter and adaptive filtering models on the same image data presented to the observers. The model decision spaces allowed for the testing of how model performance differs from observer performance.

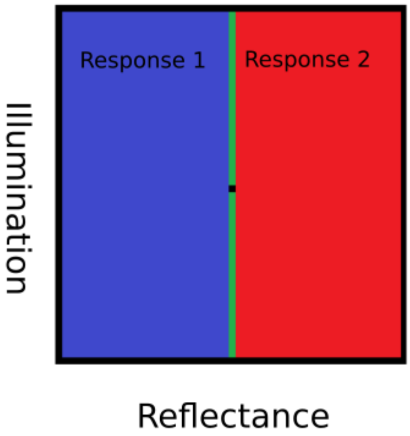


Figure 6. A hypothetical decision space. The black dot in the middle of the decision space represents the reference stimulus. Response 1 shows stimulus combinations where the observer saw the test patch as lighter than the reference patch. Response 2 shows stimulus combinations where the observer saw the reference patch as lighter than the test patch. In the decision space shown here, the observer perfectly judges the reflectance of the target with no effect from the illumination.

2 Experiment 1

The goal for Experiment 1 was to determine what stimulus properties determine people's judgements of lightness and brightness. Understanding how people perform lightness and brightness tasks requires determining what object properties regulate the visual system's decisions. From previous studies we know that lightness judgements are mostly predicted by reflectance, with some influence of luminance (Patel et al., 2018). However, the answers are much less clear when determining how brightness judgements are made. The proposed experiment aimed to accomplish two key things: testing the current evidence provided by lightness judgement studies, and establishing new evidence for how people perform brightness judgements.

Experiment 1 will use decision spaces to demonstrate observer lightness and brightness judgement data by presenting observer response for each combination of reflectance and illumination. A decision space shows the probability of observer responses as a function of both reflectance and illumination (Figure 6). The decision space is organized with the reflectance and illumination on the two axes. The judgements for lightness and brightness for each combination of reflectance and illumination are plotted separately on their respective decision spaces. The lightness and brightness judgements are based on a 2 alternative forced choice (2AFC) design. The observers were presented with a reference stimulus, and a target stimulus that is a randomly selected combination of reflectance and illumination. In figure 6, the observer is perfectly lightness constant. In other words, they are able to perfectly judge reflectance levels at all illumination levels. Response 1 represents the when the observer judged the target reflectance to be higher than the reference stimulus and response 2 is when they judged it to be lower. The line separating the two response regions indicate when observer judgements change from response 1 to response 2. Since luminance is a product of reflectance and illumination, changes in both allow for the measurement of how observer judgements change with respect to luminance as well.

In Experiment 1, the observers were presented with one of three reference patches on each trial. The experiment was be repeated with the same stimuli for both the lightness and brightness judgement tasks.

The experiment also tested the ODOG, highpass, retinex and adaptive filtering models with the same stimuli. Their results were compared with observer data and provide insights into the strategies used by observers and the models. These decision space comparisons provided a new way of testing how well current models mimic human lightness and brightness estimates.

2.1 Methods

2.1.1 Observers. The study tested two observers from York University. Both observers knew the hypotheses being tested and one was the author. The experiment has been approved by the York University Office of Research Ethics. Due to the current COVID-19 pandemic the number of observers allowed for participation was severely limited. As such, the study only had two observers. The observers were 24 to 28 years old, one male and one female. All observers provided written consent before the experiment. The observers received monetary compensation for their participation in the study.

2.1.2 Stimuli. Observers viewed a physical panel built to have a grey backdrop with two holes cut out of it where the target and reflectance patches were presented (Figure 7. The backdrop was a 61.0 cm by 30.5 cm rectangle, with the patches presented 4.5 cm horizontally on either side of center point of the backdrop. The holes had a diameter of 2.5 cm. The backdrop had a pattern of circles and rectangles of different reflectance values printed on its surface (Figure 7). These circles had a diameter between 2.0 cm and 3.5 cm and the rectangles spanned sizes from 1.0 by 2.0 cm to 2.0 by 2.5 cm. All shapes were separate with no overlap and the rectangles had two different orientations, horizontal and vertical. The grey backdrop on the panel had a reflectance of 0.24, and the shapes on the backdrop spanned a reflectance range from 0.06 to 0.77.

The target and reference patch were discs of paper transitioning smoothly from white to black. The discs were attached to two wheels placed behind the backdrop. The two circular holes in the backdrop were 2.54 cm in diameter. These discs continuously covered the range of reflectance for this experiment. The reflectance strips were visible through these holes and were presented to the observer as discs of reflectance (Figure 7). The wheel was computer controlled to rotate to the desired reflectance location on the discs behind the backdrop on each trial.

The scene had an illumination boundary at the center line of the panel. The illumination over the stimulus was controlled by a Hitachi CP-EX252N projector. The projector presented an image with two sections, one with the illumination of the reference side and the other with the illumination of the test side. The lighting present in the experiment room consisted of ambient light produced by the diffusely covered Phillips Hue bulbs and the illumination provided by the projector on the stimuli and the backdrop. The projector did not cover the entire room, while the diffuse light covered the entire room.

Observer viewing distance from the panel was 74.5 cm, with the panel covering 44.5° of visual angle and the patches covering 1.95° of visual angle. The stimulus was placed on a table with a chin rest attached to the observer side (Figure 7). The room had objects located behind the stimulus to provide the observers with additional cues to lighting within the room.

2.1.3 Procedure.

Human Observers. The experiment was presented on the physical stimulus shown in Figure 7. This apparatus was used for both the lightness and brightness judgement conditions. For both conditions the observers were presented with two patches, one reference patch and one test patch. The reference patch always showed one of the three reference reflectance and illumination combination. The reference reflectance values for the experiment were 0.10, 0.24 and, 0.48 and the reference illumination values for the experiment were 251.0π, 192.0π, 147.0π lx. The test patch had a random combination of reflectance and illumination from any point in the decision space. The patches were randomly presented on either side of the apparatus to make sure that the observer would not know one side to always be the reference stimuli. In the first condition, observers were asked to provide their lightness judgements for each trial and in the second condition, to provide their brightness judgements.

The reference patch had three values of reflectance and illumination. The lightness condition consisted of 1800 trials, 5 per combination of reflectance and illumination. These were split equally between the three reference settings over 5 repetitions of the experiment. On each trial, the observer was presented with the reference and test patches. Observer responses were based on keyboard presses. The "1" and "2" keys indicated which patch the observer reported as having a lighter shade of grey paper. The reflectance and illuminance of the test patch was randomly set from 10 values, 5 above and 5 below the reference combinations. The 10 values were set to change by steps of 8% above and below the reference stimulus. Therefore, the test combinations are spread logarithmically over the decision space, with the reference combination being the center square of the decision space. The experiment never presented observers with the reference combination at both patch locations as this would force the observer to make a judgement between two patches that are exactly the same. All trial configurations were generated and presented in a pseudo-random order with each reference and test combination appearing one time for each run of the experiment. The illumination presentations were also randomized in accordance with where the reference and test patch so that the reference illumination would always be on the reference side. Between trials the diffuse lighting in the room remained the same, while the projector lighting changed instantly between configurations. The reflectance of the reference and test patches changed gradually between trials as the discs behind the backdrop rotated to the next configuration. The trials did not have any time limit and observers could take as long as necessary for judgements. After picking a response the observers were not provided feedback.

The brightness condition stimuli were identical to the lightness condition. Observers were provided with different instructions to judge brightness. Observers were asked to pick the the disc that would require a brighter shade of grey if they were to create a painting of the stimuli. The observers were presented with the same stimuli as the lightness condition, presented in a different random order. In this condition observers were asked to indicate which of the two patches presented appeared brighter than the other.

The decision spaces were fitted with a 2D normal cumulative distribution function for Experiment 1 at all three different reference values.

For reference stimulus (r_0, l_0), this model assumes that the probability of choosing the test stimulus at point (r_i, l_i) is given by equation 4.

$$p_i = \Phi((r_i - r_0, l_i - l_0) \cdot (cos(\theta), sin(\theta)), 0, \sigma) \tag{4}$$

Here, θ controls the orientation of the normal CDF in the decision space and σ controls the speed of observer response transition from never picking the target to always picking the target. p_i denotes the probability of the observer picking the test stimulus at point (r_i, l_i). If for this test stimulus the observer chose the test patch k_i times out of n_i trials, the likelihood of the observed data given the probability p_i is given by equation 5.

$$L_i = b(k_i, n_i, p_i) \tag{5}$$

Here b(k, n, p) represents the binomial probability mass function. Assuming that the responses for all trials are independent, the negative log likelihood for the observed data is given by equation 6.

$$-l = -\sum_{j=1}^{120} \log L_i \tag{6}$$

Where j ranges from 1 to 120. Minimize equation 6 will provide the maximum likelihood estimates for the parameters θ and σ for the observer data in each decision space.

Since the θ variable determines the orientation of the fitted cdf on the decision space, it's value indicates where observers' responses switch from picking the reference to the target patch. A θ value of 0 radians represents a vertical bisector of the decision space as shown in figure 6. A θ value of 0.785 radians or 45^o degrees represents a negative diagonal bisector of the decision space. These values of θ are important benchmarks for observer decision spaces, as the vertical line represents perfect reflectance matching and the diagonal line represents perfect luminance matching.

The θ variable was then used to calculate the the Thouless ratio using equation 7.

$$T = 1 - tan(\theta) \tag{7}$$

Since the Thouless ratio values are based on theta, they provide a different way of characterizing the orientation of the 2D CDF with respect to the decision space. These Thouless ratios were calculated for each reference reflectance and illumination combination for each observer. Each decision space shows a line representing the 50% point of the 2D CDF with the orientation of the line given by Thouless ratio value. This line, hereafter referred to as the decision line, provides information about the observers' responses compared to perfect reflectance and luminance matching. The Thouless ratio for this experiment provide an indication of where observer judgements for lightness and brightness lie on a continuum from perfect brightness matching to perfect lightness matching. A Thouless ratio value of 0 represents perfect brightness matching and a value of 1 represents perfect lightness matching.

Modeling. The models used for these experiments were the ODOG, retinex, high-pass filter and contrast normalization (Land & McCann, 1971; Dakin & Bex, 2003; Blakeslee & McCourt, 1999; Shapiro & Lu, 2011). The input for each models were computer-generated versions of the stimuli used in the experiment with human observers. This allowed for the comparison of observer and model data for lightness and brightness judgements. The models responses were then used to generate decision spaces similar to the observers. The results from the ODOG and contrast normalization models were used for comparison with observer results for brightness judgements.

Figure 7. Apparatus for Experiment 1. The grey background is populated with multiple circles and rectangles with different reflectance values. The circles indicated by arrows represent the locations of the test and reference patches.

The results from the retinex and high-pass filter models were used to compare with the observer results from the lightness judgement tasks. Model responses were quantified by determining the models' responses at the patch locations. The larger of these responses was taken to be the models' judgement. This allowed for all models to be analysed in the same way as the observers, by averaging the judgements made over each combination of reflectance and illumination. Additionally, similar to the decision spaces for the observers the model decision line and Thouless ratios were calculated through the fitting process.

3 Results and Discussion

3.1 Human Observers

Overall observer Thouless ratio means for the experiment for observer JP were 0.66 and 0.50 for lightness and brightness respectively. For observer KYP the Thouless ratio means were 0.76 and 0.59 for lightness and brightness respectively. For all trials recorded, regardless of observer, the Thouless ratio means were 0.71 and 0.55 for lightness and brightness respectively. Further inspection of within subjects data between conditions shows that for both observers, at all reference conditions, the Thouless ratio values for lightness are higher than the Thouless ratio values for brightness. The difference between the conditions is not large but it is consistently present. This difference suggests that observer perception for brightness judgements is different from their perception for lightness judgements. Another interesting trend is that overall observer Thouless ratio means are above 0.50 for both lightness and brightness matching tasks. This means that observer decision lines are slightly closer to lightness matching than brightness matching, even for the brightness matching task. These findings for the lightness matching conditions for both observers are consistent with previous literature. Patel et al. (2018) found Thouless ratio values ranging from 0.67 to 0.95. Gilchrist (2006) calculated the Thouless ratios in some of Katz's classic experiments on lightness constancy, and found values ranging from 0.35 to 0.70. Both studies find that observer performance for the lightness condition are slightly below perfect lightness constancy. This is the first experiment to measure Thouless ratios for brightness judgements. As such more research is needed to make definitive conclusions especially with naive observers. For both observers the finding for this experiment suggest that people are poor at judging luminance even when specifically instructed, instead performing slightly closer to reflectance matching.

Figures 8 and 9 show the decision spaces for observers JP and KYP for Experiment 1. The decision spaces plot the observers' responses in terms of proportion of the times they choose the test patch. This allows for the visualization of the decision line, in other words, the line that defines when observers begin to switch to choosing the test patch over the reference patch. Observer data was analysed by creating a 2D matrix of the averaged response for all trials at each combination of the test reflectance and illumination. This matrix had values between 0 and 1, where 0 represented the observer always choosing the reference patch and 1 representing the observer always choosing the test patch. Therefore, lightness judgements for an observer with perfect lightness constancy would be 0 at all test locations below the reference reflectance and 1 at all test locations above the reference reflectance. These matrices were plotted into the decision spaces in Figures 8 and 9. Since there was no test patch with the same reflectance and illumination combination as the reference, the center square of the decision space had no data to plot. Therefore, for all decision spaces the center square representing the reference combination is made white.

The data from all conditions for both observers was bootstrapped to determine the confidence intervals and standard error of the fitted parameters θ, Thouless ratios, and σ. The bootstrapping consisted of 1000 iterations of non-parametric data randomly generated by selecting trials from the original dataset with replacement. The estimate of the standard error was calculated using all bootstrapped θ values for all three reflectance values for the lightness and brightness conditions for both observers. Table 1 catalogs all Thouless ratios and θ for observers JP and KYP for Experiment 1 for the lightness and brightness judgement conditions. The table also shows the 95%

confidence intervals for each Thouless ratio value and the standard errors for the θ values for both conditions and both observers in Experiment 1.

These values for the Thouless ratios provide an important benchmark for understanding observer responses. Values for the lightness and brightness conditions should show significant differences if they are different fundamental perceptual dimensions. One way to test for this is to use the z-test calculate the z-scores for the θ values using the θ estimates and the bootstrapped standard error. This allowed for significance testing between the brightness and lightness conditions within subjects. Based on the current literature on lightness and brightness perception the expectation would be that the θ values between conditions within subjects will be significantly different. Formula 8 shows how the z-scores were calculated. Here, θ_1 and θ_2 represent the the two *theta* parameters being compared and σ_1 and σ_2 represent the bootstrapped standard error for the respective θ values. Table 2 shows the significance test scores for lightness versus brightness judgements in the physical apparatus.

$$z = \frac{\theta_1 - \theta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{8}$$

Observer JP had a significant difference between lightness and brightness θ values for the 0.10 and 0.24 reference reflectance levels. Observer KYP had a significant difference between lightness and brightness judgements for the 0.24 and 0.48 reference reflectance levels. Based on the results for the two observers there is some evidence that observers have the ability to perform accurate brightness matching in Experiment 1. However, an important trend in the data of both observers is that the brightness matching Thouless ratios were halfway between perfect lightness and brightness matching. This is important to note as observer results for the lightness matching tasks was decidedly closer to the Thouless ratio for

		Observer	JP		KYP	
Experiment	Condition	Reference Reflectance	Thouless Ratio [CI 95%]	$\theta[\sigma]$	Thouless Ratio [CI 95%]	$\theta[\sigma]$
Flat-Panel	Lightness	0.10 0.24 0.48	0.457 [0.382, 0.528]	0.50 [0.09]	0.707 [0.631, 0.787]	0.28 [0.10]
			0.289 [0.209, 0.371]	0.62 [0.09]	0.750 [0.673, 0.823]	0.25 [0.11]
			0.353 [0.289, 0.371]	0.57 [0.09]	0.580 [0.500, 0.661]	0.40 [0.10]
	Brightness		-25.398 [-187.483 , 132.371]	1.53 [0.37]	0.579 [0.499, 0.657]	0.40 [0.11]
			-0.157 [-0.482, 0.090]	0.86 [0.26]	0.767 [0.69, 0.843]	0.23 [0.11]
			0.251 [0.181, 0.326]	0.64 [0.08]	0.418 [0.344, 0.491]	0.53 [0.10]
Physical	Lightness	0.10 0.24 0.48	0.729 [0.657, 0.811]	0.26 [0.11]	0.672 [0.552, 0.783]	0.32 [0.18]
			0.742 [0.675, 0.815]	0.25 [0.10]	0.862 [0.783, 0.931]	0.14 [0.10]
			0.503 [0.426, 0.582]	0.46 [0.10]	0.742 [0.655, 0.825]	0.25 [0.10]
	Brightness		0.491 [0.428, 0.553]	0.47 [0.07]	0.577 [0.485, 0.669]	0.40 [0.15]
			0.615 [0.546, 0.682]	0.37 [0.08]	0.628 [0.545, 0.712]	0.36 [0.12]
			0.391 [0.311, 0.472]	0.55 [0.11]	0.572 [0.487, 0.666]	0.40 [0.13]

Table 1

Data table showing the Thouless ratio values and 95% CI for the observers. The values recorded are for observers JP and KYP, for the flat-panel experiment's lightness and brightness tasks and the physical experiment's lightness and brightness tasks. The Thouless ratio values are recorded for each reference reflectance from top to bottom corresponding to 0.10 to 0.48 perfect lightness matching. This discrepancy between the lightness matching and brightness matching results may indicate that brightness is not a perceptual dimension with the same robustness as lightness. It is important to note that these are results from practised observers and as such their responses might not be representative of naive observers.

Conditions	Reference Values	z-scores JP	z-scores KYP
Physical Lightness vs Brightness	0.10	3.97*	1.10
	0.24	2.26*	3.86*
	0.48	1.53	-1.99*
Flat-panel Lightness vs Brightness	0.10	0.003	2.03*
	0.24	1.70	-0.30
	0.48	1.42	2.40*
Brightness Flat-panel vs Physical	0.10	0.004	-0.02
	0.24	3.51*	-2.22*
	0.48	1.98*	2.19*
Lightness Flat-panel vs Physical	0.10	4.27*	-0.44
	0.24	6.62*	2.06*
	0.48	2.24*	-2.17*

Table 2

Significance testing using z-scores within experiment 1 and 2 and between experiment 1 and 2. The z-critical value used is the $z = 1.96$ *for* $\alpha = 0.05$. *All values with asterisks indicate significantly different results.*

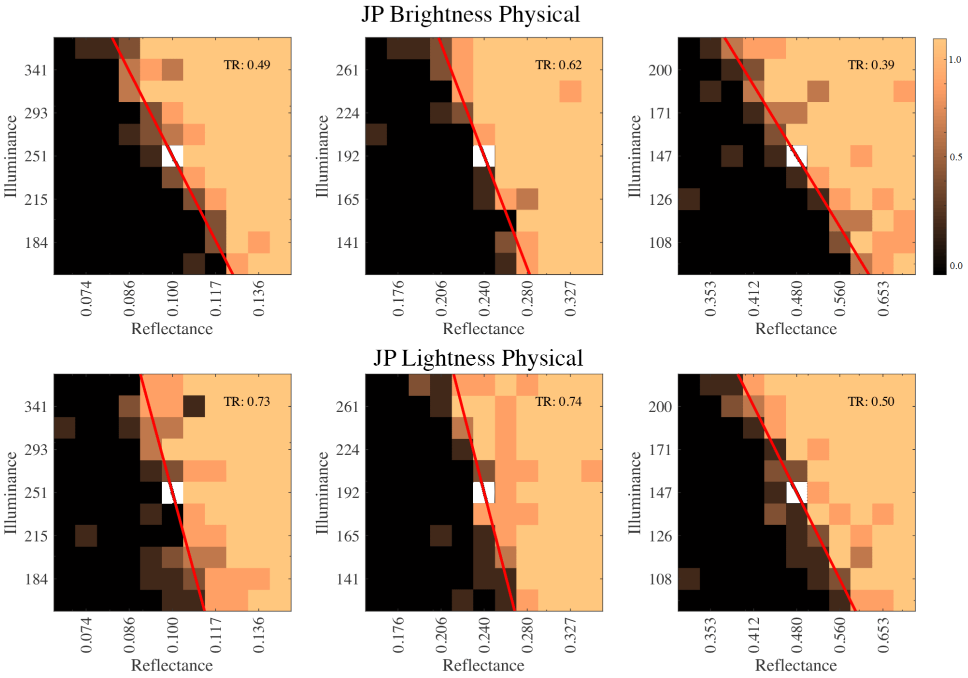


Figure 8. The brightness and lightness matching decision spaces for observer JP from Experiment 1. The TR values are the Thouless ratio values for each decision space. These are cataloged in Table 1.

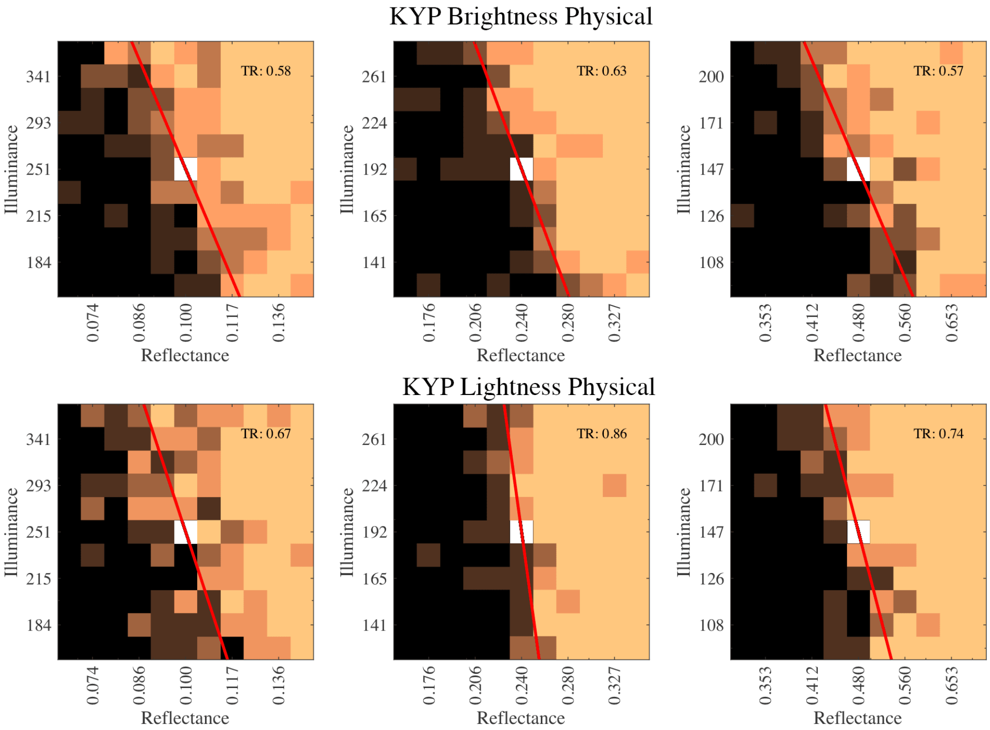
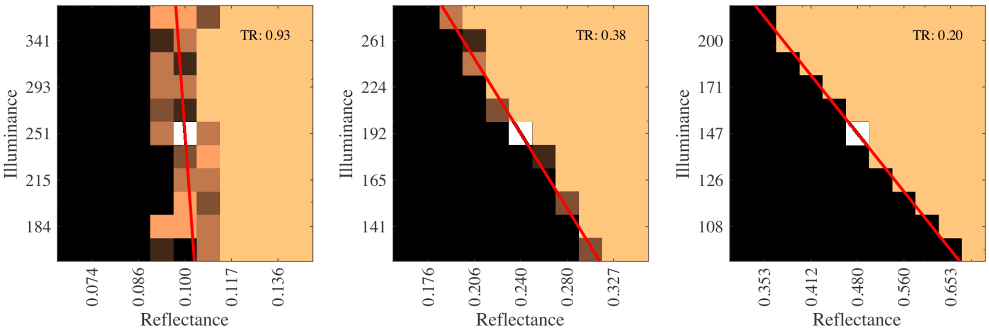


Figure 9. The brightness and lightness matching decision spaces for observer KYP from Experiment 1. The TR values are the Thouless ratio values for each decision space. These are cataloged in Table 1.

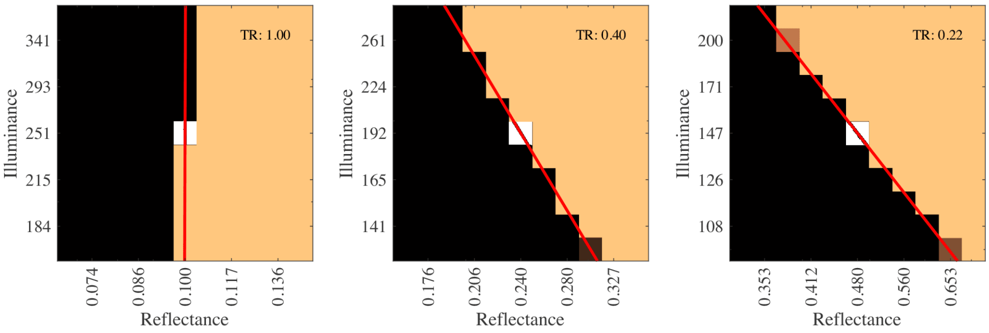
3.2 Modelling results and discussion

Figures 10 and 11 show the decision spaces for the ODOG, retinex, highpass and adaptive filtering models. The Thouless ratios estimates for the models are presented in Table 3. As a reminder, the ODOG and highpass models are brightness judgement models and as such should have Thouless ratio values similar to observer Thouless ratios for the brightness task. The retinex and adaptive filtering model are lightness judgement models and should have Thouless ratio values close to one . Additionally, the models should have internal consistency when judging the images, meaning that their results should not vary dramatically between different reflectance values. However, as seen in the figures and the table, neither of these is the case. Instead, the ODOG, retinex and adaptive filtering models appear to be matching reflectance for the lowest reference reflectance and then switch to matching almost luminance for the rest. While highpass, does negative luminance matching for the first reference reflectance, reflectance matching for the second reference reflectance and luminance matching for the third reference reflectance. Here, negative luminance refers to the model judging low illumination, high reflectance patches as being brighter than high illumination, high reflectance patches. Both retinex models have comparable results for the first and second reflectance. However the 1999 model has very noisy data for the third reflectance while the 1986 model as almost no noise. Importantly, all models have a strong trend with Thouless ratio values declining with an increase in reference reflectance. In other words, the models are getting closer to luminance matching the higher the reference reflectance. This is curiously also the trend followed by the lightness matching models as well. There does not seem to be an obvious reason for this trend to appear in the models and no such trend is present in the observer data. Based on these findings and the models' response being inexplicably and strongly tied to reference reflectance it seems the models fail at performing this simple task. Furthermore, the models also fail at predicting observer performance for this task and are more similar to each other, even though some are lightness judgement models while others are brightness judgement models.

Adaptive Filtering Model



ODOG



Highpass

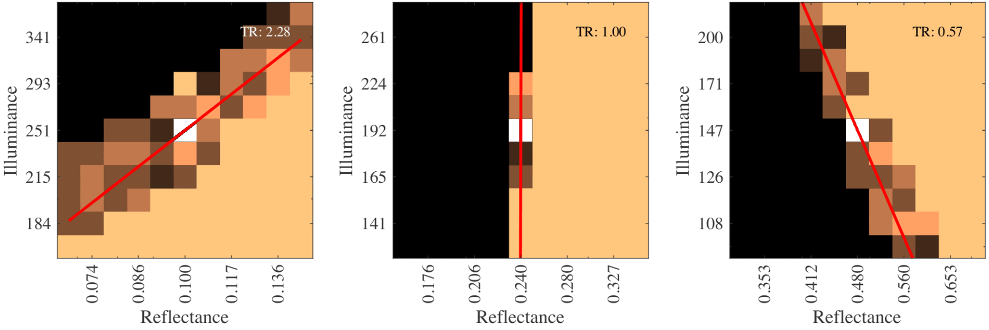


Figure 10. Decision spaces for the Adaptive filtering, ODOG and Highpass models to computer generated versions of the stimuli used in Experiment 1

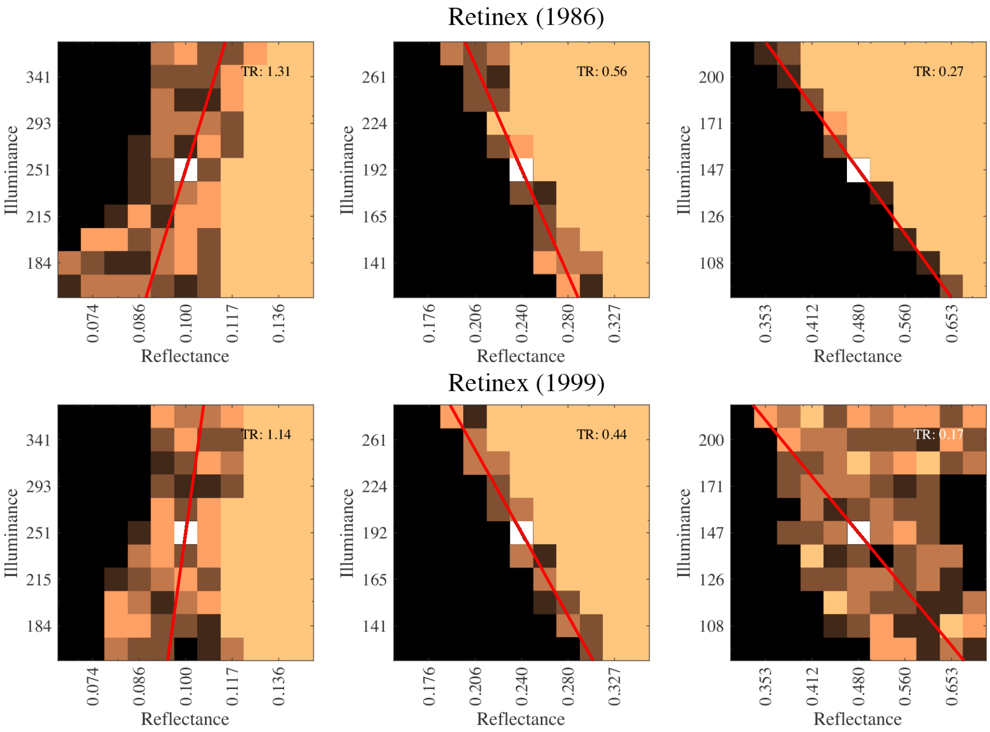


Figure 11. Decision spaces for the retinex models to computer generated versions of the stimuli used in Experiment 1

Reference Reflectance	**0.10**	**0.24**	**0.48**
Adaptive Filtering	0.93	0.38	0.20
ODOG	1.00	0.40	0.22
Highpass	2.28	1.00	0.56
Retinex 1986	1.31	0.56	0.27
Retinex 1999	1.14	0.44	0.17

Table 3

Model Thouless ratio values for each reference reflectance.

4 Experiment Two

Experiment 2 was a recreation of Experiment 1 with flat-panel stimuli instead of real surfaces and lights. This allowed us to study whether the change in presentation media had an effect on observer responses for the lightness and brightness judgements.

Some previous literature studying lightness and brightness judgements has relied on the use of flat-panel monitors (Blakeslee & McCourt, 1999; Arend & Spehar, 1993). This presents a potential problem as the stimuli presented do not have a true reflectance, only a luminance value that is displayed by the screen. For example, a real apple will have a fixed reflectance value regardless of other factors. That is, no matter the illumination in the environment, the apple will reflect a fixed proportion of light. This is in stark contrast to computer presented scenes. Here, changing the "reflectance" of the apple requires changing their luminance according to a theoretical model used to render how its surface would appear under lights. Nothing in the computer-generated stimulus actually has the intended reflectance of the apple. We rely on observers to make an interpretation of reflectance from a picture that is actually a glowing surface. Previous studies have also found that performance on lightness and brightness judgements is susceptible to how the observers are instructed to perform the task (Arend & Spehar, 1993). Additionally, there is a limited number of studies investigating differences between lightness and brightness judgements and even fewer that compare these differences in a physical and virtual environment (Blakeslee et al., 2008)

The main aim of Experiment 2 was to examine observer brightness and lightness judgements in flat-panel presentation media. The experiment provided a deeper understanding of how flat-panel displays affect the cues observers use to make lightness and brightness judgements as compared to the physical setting. Studying these topics as well as comparing how the results of the flat-screen version match up with the physical version allowed for a better picture of whether flat-panel displays provide a sufficient level of fidelity for use as replacements for real physical stimuli.

4.1 Methods

4.1.1 Observers. The same observers from Experiment 1 participated in Experiment 2. The experiment has been approved by the York University Office of Research Ethics. The observers were 24 to 28 years old, one male and one female. All observers provided written consent before the experiment. The observers received monetary compensation for their participation in the study.

4.1.2 Stimuli. This experiment used a flat-panel display to present a replication of the physical experiment, created in MATLAB. The flat-panel scene was presented on a 27 inch Retina 5K imac monitor. The physical backdrop was replaced with the monitor that displayed the stimuli. This version of the experiment was performed in an empty room with none of the additional materials seen behind the backdrop in Figure 7. The ratio of stimulus size between the physical and flat-panel presentation methods was preserved, meaning that since the horizontal length of the stimuli changed from 61.0 cm to 59.0 cm this change in ratio was also applied to all other aspects of the stimuli. Therefore, the flat-screen stimuli had dimensions of 59.0 by 29.5 cm and was gamma corrected. The grey backdrop was an image of the one used in Experiment 1 scaled down to fit the aforementioned presentation size. Hence, the circles and rectangles present on the backdrop were also adjusted to the new presentation size. This was done to keep the visual angle constant between Experiment 1 and Experiment 2.

The trials were randomized but still had the same parameters as the ones presented in Experiment 1, allowing for consistency between the two decision spaces. Using a photometer, the reflectance values for all patches in Experiment 2 were measured to be the same as the ones presented in Experiment 1. All patches presented were measured to have the same luminance as the patches presented in Experiment 1. The illumination boundary was still presented in the center of stimuli and the patches were 4.3 cm from the center line and 8.6 cm from each other. The presentation patches were 1.93 cm in diameter and randomly switched between presenting the target and reference reflectance. Additionally, the illumination also switched randomly, but in correspondence with the reflectance between the two presentation patches. A stimulus trial from the flat-panel experiment is shown in Figure 12.

4.1.3 Procedure. In the flat-panel condition, observers used a chin rest to view the experiment presented on the monitor. The location of the chin-rest was unaltered from Experiment 1

The conditions in Experiment 2 included brightness and lightness judgements just like in Experiment 1. Observers were presented with the same target and reference patch reflectance values as in Experiment 1. The range of illuminations was also the same as in Experiment 1.

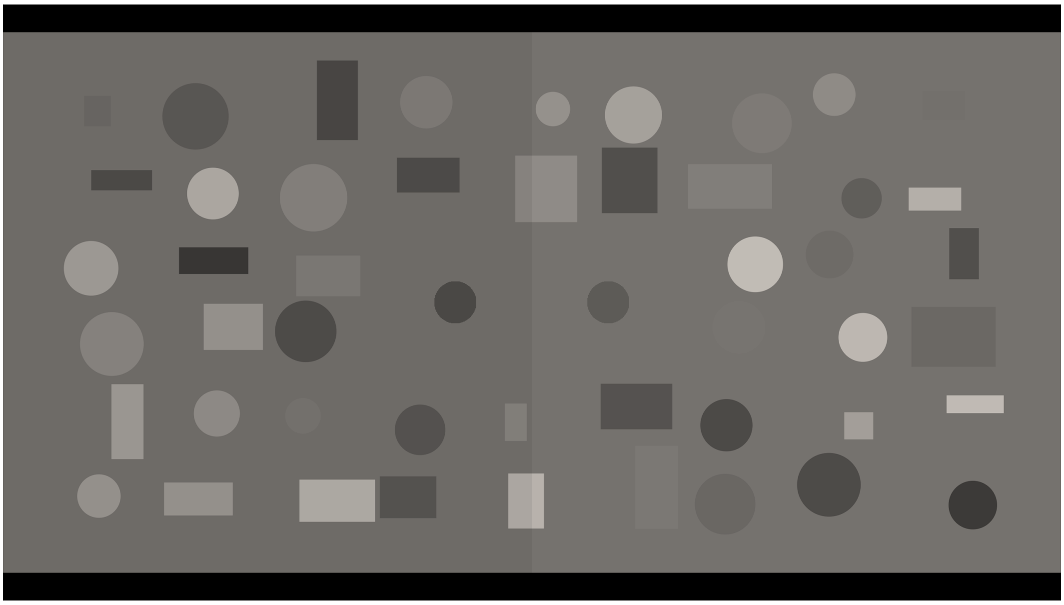


Figure 12. Flat screen version of the lightness and brightness experiment

5 Results and Discussion

5.1 Analysis

Figures 13 and 14 show the decision space data for observers JP and KYP for the flat-panel experiment. The same analysis from Experiment 1 was repeated for Experiment 2. Observer data was fitted with the 2D cdf and the resultant decision line plotted on the decision spaces. The Thouless values were, again, calculated by monotonically transforming the θ variable from the minimized error function used to generate the cdfs.

Table 1 catalogs all Thouless ratios for observers JP and KYP for Experiment 1 for the lightness and brightness judgement conditions. The table also shows the 95% confidence interval for each Thouless ratio value for both conditions in both versions of the Experiments.

Experiment 2 used the same significance testing as Experiment 1. Table 2 shows the results for the significance testing between lightness and brightness judgements in flat-panel, as well as the difference between presentation methods within conditions. Observer JP had no significant difference between the lightness and brightness judgements for flat-panel presentation. Observer KP had a significant difference for the 0.10 and 0.48 reflectance values. Based on these findings there is some evidence that observers are able to accurately perform lightness and brightness when presented on a flat-panel display. Although for the brightness task, it is important to note that performance is quite far from perfect brightness matching.

Comparing the results from Experiment 1 and Experiment 2 within conditions shows how observer responses change based on the presentation method. For the brightness condition, both observers had a significant difference between responses for the physical and flat-panel presentation at the 0.24 and 0.48 reflectance levels. Furthermore, for the lightness condition, observer JP had a significant difference at all reflectance levels between physical and flat-panel responses and observer KP had a significant difference at the 0.24 and 0.48 reflectance levels between the physical and flat-panel responses. Based on these preliminary results, there is evidence that observer responses are significantly different based on the presentation media chosen for the experiment.

Overall observer Thouless ratio means for the flat-panel version of the experiment are quite different from the observer responses from the physical version. Thouless ratio values for observer JP were 0.37 for the lightness condition and -8.33 (excluding the first reference reflectance outlier, 0.21) for the brightness condition. Thouless ratio values for observer KYP were 0.68 for the lightness condition and 0.59 for the brightness condition. For all trials recorded, regardless of observer, the Thouless ratio means were 0.52 for the lightness conditions and -3.87 (without the outlier 0.43) for the brightness condition. Between both observers five out of 6 conditions had lightness Thouless ratio values greater than brightness Thouless ratio values. The differences between conditions in this experiment were even lower than the differences in the physical experiment. This could mean that even practised, expert observers had a harder time performing the brightness task in the flat panel experiment. Observer lightness Thouless ratio means for the flat-panel experiment are lower than means in the physical experiment. This result is consistent with previous evidence showing lower TR values for flat-panel presentation than physical presentation (Patel et al., 2018). The decision space for the first reference combination in the brightness task for observer JP shows an instance where the Thouless ratio is below zero. Here the observers seems to be performing illumination matching. This along with the large variability in responses indicates that it is likely the observer was unable to accurately perform brightness matching around this reference combination. Instead, the results for this reference combination show that the observer was performing illumination matching. Overall brightness Thouless ratio means for the flat-panel experiment are also smaller than the means in the physical experiment. Importantly, they are still quite far from the value of 0 that would be expected for perfect luminance matching. For observer KYP, the more practiced observer the Thouless ratio mean for the brightness condition is closer to reflectance matching. For observer JP, the less practiced observer the task was substantially more difficult in the brightness condition as seen in their results for the first reference stimulus where they are performing illumination matching. Without this outlier, observer performance was similar to the physical experiment, but with higher variability. Based on these results lightness constancy for both observers seems to be worse in the flat-panel experiment compared to the physical experiment. Brightness matching, although better (with the exception of the outlier), still has Thouless ratio means much greater than zero. For both observers the findings for this experiment suggest that people are poor at judging luminance even when specifically instructed, instead performing slightly closer to reflectance matching.

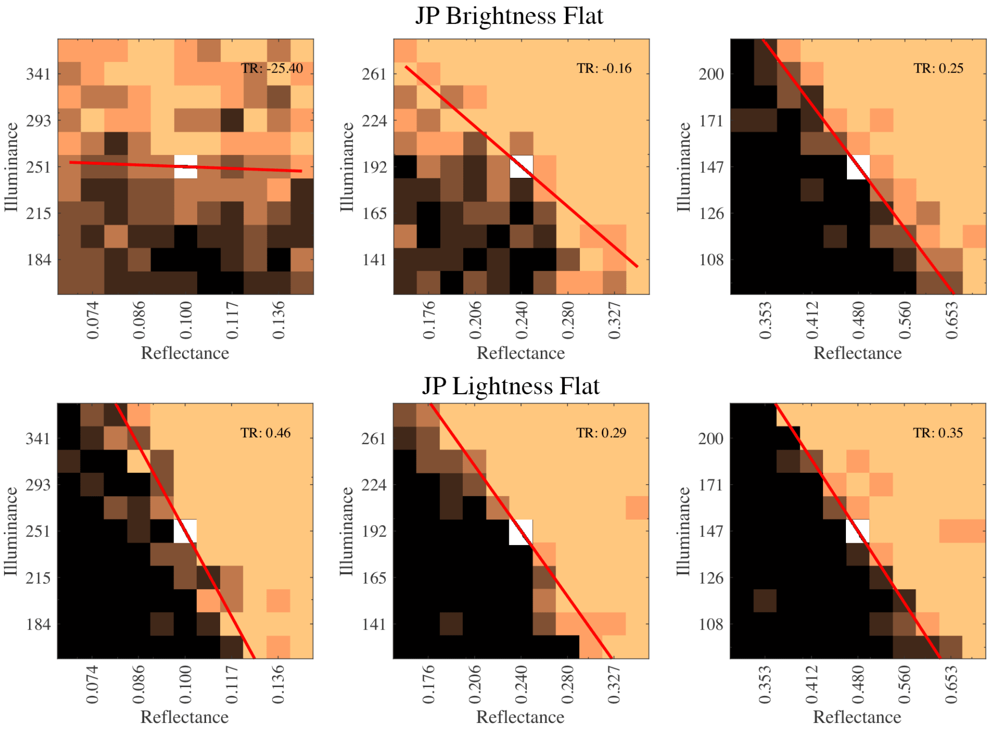


Figure 13. Decision spaces for the brightness and lightness condition for observer JP for the flat-panel experiment.

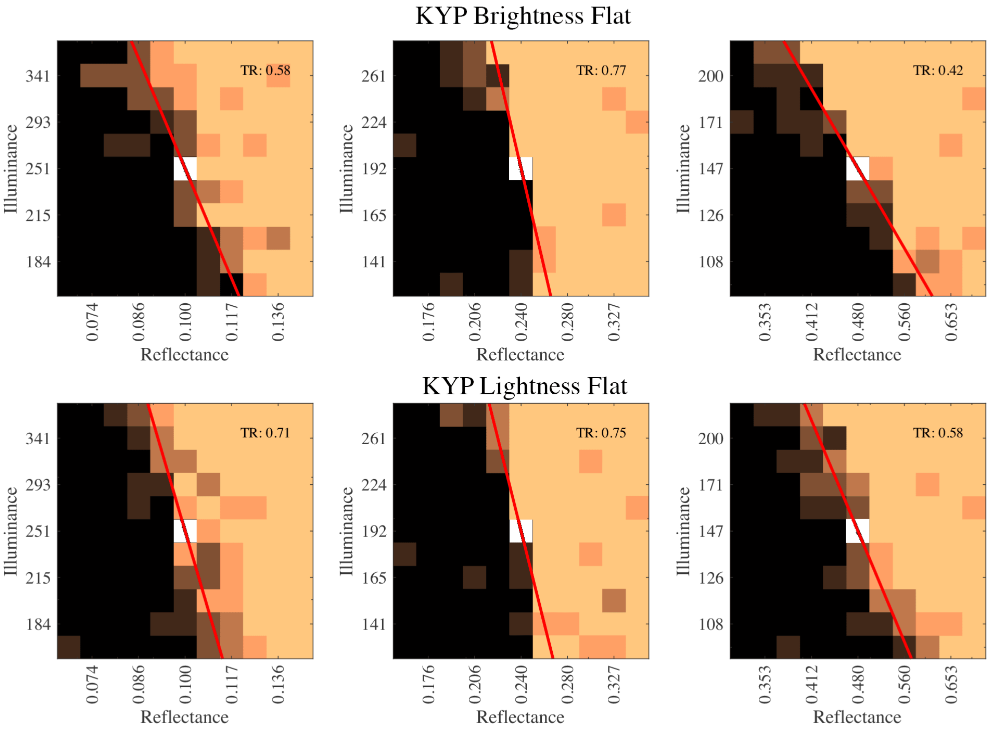


Figure 14. Decision spaces for the brightness and lightness condition for observer KYP for the flat-panel experiment.

6 Discussion

6.1 Summary of Findings

Experiment 1 contributes to the lightness literature by presenting first time Thouless ratio measurements of brightness perception. Furthermore, Experiment 1 tested the current established lightness and brightness perception models' performance on the task on two important basis. Firstly, testing whether these perception models actually perform lightness and brightness perception and secondly, for the first time testing whether the models represent how people perceive lightness and brightness. The results from Experiment 1 provide evidence for people's ability to accurately judge reflectance over a wide range of reflectance and illumination, confirming previous findings in the literature. There is preliminary evidence that observers responses for the brightness condition are different from the lightness condition. However, the Thouless ratios calculated for observer brightness judgements are much larger than what would be expected for luminance matching. Taking into account that the observers in this study were highly practised, expert observers, it seems that brightness perception is a much more difficult task than lightness perception. Additionally, observers' Thouless ratios show some variability between reference values for the lightness and brightness conditions. This variability is not unexpected as previous studies have found much larger individual differences Murray (2020); Kim et al. (2018). Further research needs to be conducted with naive observers as their response might be quite different from the observers.

Experiment 1 tested the ODOG, highpass, adaptive filtering and retinex models. The model responses to the stimuli do not generate the expected results. The lightness perception models, retinex and adaptive filtering, as well as the brightness filtering model ODOG perform similar to each other. Curiously, each model has an completely different decision lines based on the reference reflectance, meaning that the models are not consistent over the spectrum of reflectances. The highpass model performs a combination of luminance matching, reflectance matching and negative luminance matching for the three reference reflectances. The models also fail at predicting observer behaviour on either presentation method as the observers show no such trend between reference reflectances. Further testing should be done to investigate why these models have such viscerally different responses over the range of reflectances and why brightness and lightness perception models have the same responses for such a simple task.

Experiment 2 examined the importance of presentation media on lightness and brightness perception. It aimed to fill the current gap in the literature of testing the efficacy of using flat-panel displays in place of physical stimuli. Additionally, the experiment aimed to investigate if and how lightness and brightness perception changes based on presentation media. Experiment 2 provides evidence for a significant difference in observer lightness and brightness judgements for the same stimuli when presented on different media. Observers had significantly different results for a majority of conditions between physical and flat-panel stimuli. The experiment provides preliminary evidence that lightness and brightness are perceived to be different from each other when presented on a flat-panel display. Observer Thouless ratios were on par with previous literature for the lightness condition. However, similar to the physical experiment Thouless ratios were larger than expected for the brightness condition. Observers also showed increased variability between references in the same condition when compared to Experiment 1. These individual differences again do not seem unexpected for brightness judgements as they are consistent with previous studies showing variations across conditions Murray (2020); Kim et al. (2018).

Observer JP's results for the brightness condition fall far outside of the expected range of variability. The large variability suggest an inability to correctly perform the task, especially at the lowest reflectance. Due to an inability to properly identify differences between the presented patches for the lowest reflectance reference combination, the observer relies on illumination for their judgements. This is in stark contrast to observer KYP as they show very consistent results for both conditions. As observer KYP is the more practised observer, this seems to show that observers are able to make lightness and brightness judgements on a flat-panel similar to the physical presentation with extensive practise. More research with naive subjects is required to evaluate their performance and see whether they are able to correctly perform the task and judge luminance.

6.2 Answers to motivating questions

Achromatic surfaces have different perceptual dimensions such as lightness, gloss, surface brightness, translucency. Whether the human visual system has access to these different dimension is a matter of much debate. Previous literature has tried to test the presence of these dimensions in the visual system under different conditions and with different instructions (Arend & Spehar, 1993). One study in particular used multidimensional scaling to determine if the result correspond to any of the aforementioned perceptual dimensions (Logvinenko & Maloney, 2006). Their findings suggest the existence of two dimensions roughly corresponding to reflectance and illuminance. The theory also proposed that observers have no perceptual access to either dimensions. Based on these earlier findings, the hypothesis for the Experiment 1 was that observers would perform better on the lightness task than on the brightness task. Additionally, the lightness and brightness models were hypothesized to perform their respective judgements with near perfect accuracy. For Experiment 2 the hypothesis was the observers' results for lightness and brightness would be different from Experiment 1 because of the change in presentation methods. Additionally, the lightness perception scores were predicted to be worse than Experiment 1 as the presentation did not truly have "reflectance".

The results for Experiments 1 and 2 largely confirm the predicted hypothesis. In Experiment 1, observers performed better on the lightness judgement task compared to the brightness judgement task. Observer results for the brightness judgement task were closer to lightness perception than brightness perception. These results seem to confirm the fact the brightness perception is a much harder task for people than lightness perception, even with clear instructions. Furthermore, peoples brightness judgements are quite far from brightness matching and are closer to lightness matching. This is consistent with the idea that luminance is not a perceptual dimension like reflectance and illumination. Based on these findings, Experiment 1 shows that people are better at lightness perception than brightness perception and that luminance is likely not a perceptual dimension. In Experiment 2, observer results for both tasks were significantly different than the in Experiment 1. This shows that presentation methods have a effect on observers ability to perform lightness and brightness judgements. Specifically, stimulus presentation on a flat-panel screen causes the results for both tasks to skew towards brightness matching. The results also show more variability within each task, this might indicate that the task was more difficult for observers than the physical version.

One place where the prediction for the results failed was with the lightness and brightness models. None of the models tested performed the task correctly. The lightness and brightness models had results similar to each other rather than lightness models being different from brightness models. The general trend, regardless of whether it was a lightness or brightness perception model, was almost perfect reflectance matching for the first reference combination and in-between perfect lightness and brightness perception for the other two reference combination. Only the highpass model differed from this trend, performing negative luminance matching for the first reference combination, perfect lightness matching for the second reference combination and in-between lightness and brightness matching for the third reference combination. There is no discernible reason to explain the models' failure at such a simple task. Furthermore, all models changing strategy based on reflectance is an extremely curious effect, since their responses should remain the same across the entire reflectance continuum.

6.3 Limitations and Future Work

One of the biggest limitations for this study was the lack of participants. Due to the COVID-19 pandemic, only two participants ran in the study. Furthermore, both participants were expert observers and aware of the purpose of the study. More testing needs to be done with naive observers to investigate whether the difference in Thouless ratios between the lightness and brightness conditions originate from the observers being practices. Based on the results of this experiment there is evidence that flat-panel presentation of physical stimuli has an effect on observer responses. Therefore, previous studies investigating lightness and brightness judgements conducted with flat-panel presentation should be redone and examined for their validity. Additionally, more research needs to be conducted on the continued use of flat-panel displays as a viable medium for lightness and brightness judgement experiments, especially with naive observers.

Future research based on the work presented here could test the efficacy of virtual reality (VR) as a replacement for physical stimuli for lightness and brightness judgement tasks. There are some distinct advantages of using VR as a method for measuring properties of lightness and brightness perception. It allows for precise stimulus control that might otherwise be difficult or impossible to manipulate in the real world. The ability to test and control for these factors individually allows for a better understanding of what role they play in visual perception. VR also allows for an immersive environment with 3D stimuli and covering observers' visual field completely. These advantages over flat-panel presentation might make VR a more analogous presentation method to physical stimuli than flat-panel screens.

References

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*(5142), 2042–2044.

Arend, L. E. & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. illuminance variation. *Perception & Psychophysics*, *54*(4), 446–456.

Blakeslee, B. & McCourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. *Vision Research*, *39*(26), 4361–4377.

Blakeslee, B., Reetz, D., & McCourt, M. E. (2008). Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments. *Journal of Vision*, *8*(11), 3–3.

Bloj, M., Kersten, D., & Hurlbert, A. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, *402*(6764), 877–879.

Boyaci, H., Doerschner, K., & Maloney, L. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *Journal of Vision*, *4*(9), 1–1.

Boyaci, H., Doerschner, K., & Maloney, L. (2006). Cues to an equivalent lighting model. *Journal of Vision*, *6*(2), 2–2.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, *3*(8), 2–2.

Brainard, D. H. & Hurlbert, A. C. (2015). Colour vision: understanding #thedress. *Current Biology*, *25*(13), R551–R554.

Dakin, S. C. & Bex, P. J. (2003). Natural image statistics mediate brightness 'filling in'. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1531), 2341–2348.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Frankle, J. & McCann, J. (1983). Method and apparatus for lightness imaging: Usa.

Gilchrist, A. (2006). *Seeing black and white.* Number 40. OUP USA.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., & Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*(4), 795.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing-and test-field luminances. *Journal of experimental psychology*, *50*(2), 89.

Kim, M., Gold, J. M., & Murray, R. F. (2018). What image features guide lightness perception? *Journal of vision*, *18*(13), 1–1.

Land, E. H. & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, *61*(1), 1–11.

Logvinenko, A. D. & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics*, *68*(1), 76–83.

McCann, J. (1999). Lessons learned from mondrians applied to real images and color gamuts. In *Color and Imaging Conference*, volume 1999, (pp. 1–8). Society for Imaging Science and Technology.

Menshikova, G., Bayakovski, Y., Luniakova, E., Pestun, M., & Zakharkin, D. (2013). Virtual reality technology for the visual perception study. In *Transactions on Computational Science XIX* (pp. 107–116). Springer.

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision*, *14*(9), 15–15.

Murray, R. F. (2020). A model of lightness perception guided by probabilistic assumptions about lighting and reflectance. *Journal of Vision*, *20*(7), 28–28.

Murray, R. F. (2021). Lightness perception in complex scenes. *Annual Review of Vision Science*, *7*.

Patel, K., Munasinghe, A., & Murray, R. (2018). Lightness matching and perceptual similarity. *Journal of Vision*, *18*(5), 1–1.

Pritchett, L. M. & Murray, R. F. (2015). Classification images reveal decision variables and strategies in forced choice tasks. *Proceedings of the National Academy of Sciences*, *112*(23), 7321–7326.

Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S., & Brainard, D. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision*, *4*(9), 7–7.

Shapiro, A. & Lu, Z.-L. (2011). Relative brightness in natural images can be accounted for by removing blurry content. *Psychological Science*, *22*(11), 1452–1459.

Snyder, J. L., Doerschner, K., & Maloney, L. (2005). Illumination estimation in three-dimensional scenes with and without specular cues. *Journal of Vision*, *5*(10), 8–8.